

Advanced Machine Learning
Class Test II
Time: 1 hour. Marks: 30

1. Let a univariate random variable x be distributed according to a mixture of two Gaussian. The parameters (mixture weight, mean, variance) of the Gaussian are $(w_1 = 0.6, \mu_1 = -1, \sigma_1 = 2)$ and $(w_2 = 0.4, \mu_2 = 1, \sigma_2 = 3)$. We want to generate samples from this mixture distribution using the Metropolis-Hastings sampling algorithm. The proposal distribution used is also a Gaussian $\mathcal{N}(x, 0.1)$, that is centred at the current sample x and has a variance 0.1.

We list five random numbers generated from the zero-mean unit-variance Gaussian distribution $\mathcal{N}(0, 1)$: $\{0.35, 0.05, 0.87, 0.62, -0.69\}$. Also, five uniform random numbers in the range $[0, 1]$ are as follows: $\{0.81, 0.09, 0.15, 0.14, 0.65\}$.

Let the initial sample be $x_0 = 1$. Obtain the next two samples x_1, x_2 using the Metropolis-Hastings algorithm. Show the steps clearly.

[15]

2. A corpus has 3 documents and 5 words {hall, class, exam, recall, campus} and 2 topics. We want to perform Latent Dirichlet Allocation on the corpus. At certain iteration of the Gibbs sampling the documents have the following topic distribution:

Doc1: $recall^1, exam^1, campus^1, exam^1, recall^1, recall^1, campus^1$
Doc2: $recall^1, exam^1, exam^2, hall^2, campus^1, class^2, exam^1, recall^1$
Doc3: $hall^2, exam^2, class^2, exam^2, hall^2, hall^2, class^2, exam^2$

The superscripts 1 and 2 denote if currently the word is generated from Topic 1 or Topic 2. The topic-document Dirichlet distribution has parameter $\alpha = 0.1$, and the word-topic Dirichlet distribution has parameter $\eta = 0.3$.

(a) Using the above data, estimate the parameter $\beta_{recall}^1, \beta_{recall}^2$, representing the probabilities of the word ‘recall’ being generated from each of the topics.

(b) Estimate the topic proportion $\theta_{Topic1}^{Doc2}, \theta_{Topic2}^{Doc2}$ for Doc2.

(c) In the next iteration of Gibbs sampling what topic would be assigned to the first word of Doc1. Show the steps.

[5+5+5]